

## Association for Information Systems AIS Electronic Library (AISeL)

---

AMCIS 2009 Proceedings

Americas Conference on Information Systems  
(AMCIS)

---

2009

# Sloppy Tags and Metacrap? Quality of User Contributed Tags in Collaborative Social Tagging Systems

Hongwei Zhu

*Old Dominion University, hzhu@odu.edu*

Harris Wu

*Old Dominion University, hwu@odu.edu*

Follow this and additional works at: <http://aisel.aisnet.org/amcis2009>

---

### Recommended Citation

Zhu, Hongwei and Wu, Harris, "Sloppy Tags and Metacrap? Quality of User Contributed Tags in Collaborative Social Tagging Systems" (2009). *AMCIS 2009 Proceedings*. 438.

<http://aisel.aisnet.org/amcis2009/438>

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2009 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# Sloppy Tags and Metacrap? Quality of User Contributed Tags in Collaborative Social Tagging Systems

**Hongwei Zhu**

College of Business and Public Administration  
Old Dominion University  
hzhu@odu.edu

**Harris Wu**

College of Business and Public Administration  
Old Dominion University  
hwu@odu.edu

## ABSTRACT

There have been an increasing number of social tagging systems on the web that allow users to contribute tags collaboratively to describe various resources. As the number of tags continue to grow, it is important that we can describe and measure their quality. There have been conflicting opinions about the quality of user-contributed tags but the issue has not been systematically studied. In this work-in-progress paper, we propose a preliminary methodology for assessing tag quality. The methodology identifies three aspects of tag quality in a social tagging system: individual tags, collections of tags, and the association of the tags with the corresponding resources. The proposed methodology is being empirically evaluated using the tags of several social tagging systems. In addition to validating and refining the tag quality assessment methodology, the empirical analysis is expected to provide new findings about various properties of the tags. These findings will be used to improve the design of tagging systems by incorporating mechanisms that induce the contribution of high quality tags.

## Keywords

Information quality, tag quality, social tagging system, social networks.

## INTRODUCTION

A collaborative social tagging system (Hammond et al. 2005) allows a user to tag certain digital resources using a few keywords from uncontrolled vocabularies (i.e., anything a user can type in subject to certain system restrictions, e.g., one word per tag), and subsequently retrieve the items using the tags. For example, Flickr allows users to tag photos that they upload to the site, Youtube allows users tag video clips that they upload to the site, CiteULike and Connotea allow users to tag academic references, and Delicious allows users to bookmark URLs with tags. These tagging systems are collaborative and social because the resources and tags are often shared amongst the users and user groups. Such systems have been growing rapidly in the past few years. For example, in 2005, Flickr had 775,000 registered users and 19.5 million photos (Kuchinskas 2005). Within two years, these numbers went up to 5 million and 250 million, respectively (Ames and Naaman 2007).

As the number of tags continues to grow in these tagging systems and people start to use the tags for purposes beyond the intent of the tag contributors, it is important that we know the quality of the tags and understand the implications of tag quality to various applications of the tags and to the design of tagging systems.

Tags are entered using free-form text boxes, thus users can enter whatever they like. As a result, there may be certain degrees of sloppiness in the tags, such as typos, acronyms of which no one else but the tag contributor knows the meaning, arbitrary concatenation of multiple words into one (Guy and Tonkin 2006). The contributors may have certain bias and personal agenda (e.g., to allure visitors to a resource) in entering tags not relevant to the resources at all, thus the tags may be a collection of “metacrap” (Doctorow 2001). On the other hand, the user-contributed hierarchy-free tags may be more useful collectively because multiple ways of organizing information can emerge from the tags to satisfy diverse needs (Shirky 2005). The notions of folksonomies (Mathes 2004) and ethnoclassification (Merholz 2004) have directed much of the discussion to the usefulness of the tags as a replacement or complement of the formal hierarchical taxonomies created by specialists (Shirky 2005). The focal issue in folksonomy discussions is whether the tags fit for classification or categorization of the corresponding resources. When we subscribe to the “fitness for use” notion of data quality (Wang and Strong 1996), the folksonomy discussions allude to the quality of tags for purposes of organizing the resources.

How should the quality of tags be defined? How can we measure tag quality? Despite the numerous discussions mentioned earlier, no studies answered the questions directly. Prior research on data quality has discovered as many as 16 data quality

dimensions, such as accuracy, objectivity, accessibility, relevancy, value-added, and interpretability; these quality dimensions can be organized into four categories: intrinsic, contextual, representational, and accessibility (Wang and Strong 1996). The quality of the same data can be different to different stakeholders and for various purposes of using the data (Pipino et al. 2002). A survey-based method to systematically assess data quality perceived by different stakeholders (Lee et al. 2002) has been used successfully in various organizational settings. There have been studies that examine the quality of metadata in digital libraries (Guy et al. 2004; Stivia and Gasser 2008). Several recent studies on tags (Golder and Huberman 2006; Guy and Tonkin 2006; Marlow et al. 2006) have revealed certain tagging patterns (e.g., the power law distribution) and observed tag properties relevant to tag quality (e.g., the use of compound words such as *white\_cat* and *programming/C++*). While many of the data quality concepts and methods developed in prior research are relevant and applicable to tag quality, the unique social characteristics of the tagging systems warrant a dedicated in-depth study on the quality of tags.

In this work-in-progress paper, we present a preliminary methodology for assessing tag quality. We are in the process of collecting tags from several social tagging systems and will perform various statistical and data mining analyses on the collected tags. In addition to evaluating the tag quality assessment methodology, the empirical findings will help improve the design of tagging systems to induce high quality tags. Below we first review various tagging systems and provide a definition for tag quality. Then we propose the quality assessment metrics and discuss the empirical study.

### BACKGROUND: TAGGING SYSTEMS AND HOW TAGS ARE USED

Properties of tagging systems affect various aspects of the tags, including certain quality aspects of the tags. The properties, design choices of the properties, and implications of these choices are discussed in (Marlow et al. 2006). Table 1, adapted from (Marlow et al. 2006), summarizes these factors. We should note that the implications are mainly based on intuitions and casual observations.

System property	Category	Implication
Tagging rights	1. Self-tagging: tag own resources only 2. Free-for-all: tag any resource 3. Configurable permission	Nature of tags
Tagging support	1. Blind: cannot see tags of same resource by others 2. Viewable: can see tags of same resource by others 2. Suggestive: system provide tag suggestions	Convergence of tags
Tag aggregation	1. Bag: keep duplicate tags of different users for same resource 2. Set: no duplicate tags for same resource	Availability of certain statistics
Resource type	1. Textual 2. No-textual	Nature of tags
Resource source	1. User contributed: e.g., photos on Flickr 2. System supplied: e.g., Yahoo! Podcast 3. Public:, e.g., the URL's on Delicious	Nature of tags
Resource connectivity	1. Links: there are links among resources 2. Groups: resources are grouped 3. None: no link or group	Convergence of tags
User connectivity	1. Links: there are links among resources 2. Groups: resources are grouped 3. None: no link or group	Convergence of tags

**Table 1. Properties of Tagging Systems and Their Implications**

Nature of tags includes characteristics such as the number of tags for a resource, whether the tags reflect the descriptions of the resources or the opinions of the users about the resources, and the heterogeneity or richness (e.g., the use of synonyms with subtle differences) of the tags. Convergence of tags refers to the tendency that the users use the same set of tags for the resources. The bag aggregation model retains information that allows for analysis of the tagging choices by different users. Delicious uses the bag model; Flickr and Youtube use the set model.

In addition to system properties, tagging motives and purposes also affect the choices of tags. When a user's main objective is to organize the resources for retrieval by the user at a later time or for discovery by and sharing with other users, the tags tend to be descriptive and may converge to a set of commonly used tags (Marlow et al. 2006). However, when the user's main objective is to communicate opinions, the tags tend to be more subjective and the collection of tags for the same resource tend to be more heterogeneous. In addition to the differences in the motives that lead the users to contribute tags,

there are also different functions that different tags serve. An early empirical study of the tags in Delicious identifies seven different functions (Golder and Huberman 2006). Most of the functions are descriptions about the resources in terms of the topic, type (e.g., blog, book, etc.), owner or creator, and the user's opinion about the resources using adjectives such as *scary*, *funny*, or *stupid*. Self-reference tags such as *mycomments* describe the relationship between the resource and the tag contributor; task organizing tags such as *toread* allow the user to organize resources related to certain tasks and can be considered to be metadata for administrative purposes (Taylor 2003). Obviously, these different functions should be taken into consideration when assessing tag quality, e.g., administrative tags can have high quality to the contributor but low quality to other users who want to classify the tagged resources.

The tags can be used for searching and browsing purposes. Browsing is often assisted with other convenient features such as a tag cloud where the font size of a tag is proportional to the frequency of the tag being used for different resources. Figure 1 is an example of the tag cloud of the top 199 tags on Delicious on March 1, 2009 (the cloud changes with time as the users constantly tag their bookmarks using the system). The most popular tag is design (used in 7,755,095 bookmarks), the next is blog (used in 5,943,852 bookmarks), and the 199th popular is artist (used in 325,518 bookmarks). Clicking on a tag in the cloud will take the user to a page that lists all bookmarks that use this tag. A preliminary study shows that tag cloud is useful browsing tool for general (as opposed to specific) information-seeking tasks (Sinclair and Cardew-Hall 2008).

Figure 1. Cloud of Popular Tags at Delicious

In addition, tags can be used to induce ontologies and infer other relationships amongst the tags and their corresponding resources (Wu et al. 2006). These relationships can be used to further improve user's searching and browsing experience. It is also possible to identify experts on a given topic by analyzing the tags. Tags can also be used for other purposes not currently supported by existing tagging systems.

## A METHODOLOGY FOR ASSESSING TAG QUALITY

The methodology for assessing tag quality includes a definition of tag quality from a social utility perspective and a set of metrics for quality measurement.

Given that tags are contributed by users who are not subject to any authority, they can be idiosyncratic and are often created to serve peculiar purposes of tag contributors (e.g., tags such as "silly" to voice opinion and "toread" to organize personal tasks). Thus we assume that tags have utility to the contributors themselves. However, we are interested in how to make tags benefit the whole user community. In this context, we define tag quality as the degree of support for the whole user community's usage of social tagging systems, while the usage can be anything from productive knowledge work in a digital library to leisurely pastime on a social networking site.

We develop the metrics for tag quality measurement by adapting the four categories of data quality: intrinsic, representational, contextual, and accessibility quality (Wang and Strong 1996). Largely parallel to the first three categories, we identify three aspects of tag quality. We will study the accessibility quality as a future research topic, as it relates to the quality of social tagging systems such as ease of use and security, rather than tags themselves. The three aspects are:

- Intrinsic quality of tags
- Aggregate quality of a tag collection (both representational and contextual)
- Contextual quality of tag-resource association

Each aspect has its own applicable quality dimensions and their corresponding definitions. The assessment criteria may also vary according to the purposes of use (i.e., the “use” part of “fitness for use”). For example, the accuracy dimension for individual tags’ intrinsic quality is mainly determined by syntactic error-freeness (e.g., no typos according to a set of dictionaries), while accuracy for tag-resource association is mainly concerned with the degree to which the tag reflects a certain aspect of the resources. Completeness maybe defined for the tag collection aspect as the size of the collection and certain measurement of richness of terms used for the tags. Value-added dimension can be defined for tag-resource association aspect as a certain distance measure between the set of tags assigned to a resource and the metadata of the resources.

Our ongoing research is to identify the quality dimensions applicable for each aspect and develop their definitions and assessment metrics. This task will be accomplished iteratively in conjunction with the empirical analysis of tags in several tagging systems. We are developing a survey instrument that will help develop quality dimensions and assessment metrics important to users. The survey will be pre-validated with undergraduate and MBA students in a national university. The validated survey will be administered online to collect data from user communities of several social tagging systems. Based on the survey we will validate the proposed quality aspects through factor analysis. Below is an incomplete list of quality dimensions and metrics in each aspect of our tag quality assessment methodology:

#### Intrinsic quality of tags:

- Correctness: number of typos
- Unambiguity: number of polysemous words
- Conciseness: number of repeated or redundant words in tags

#### Aggregate quality of a tag collection:

- Consistency: number of synonyms and stemmed words
- Unambiguity: number of homonymous words
- Completeness: size of tag cloud
- Presentation: probabilistic distribution
- Semantic arrangeability: Observed Linguistic Precoordination (OLP, a library science term), Compositionality
- Sufficiency: Size of tag collection versus number of resources and number of users, usage probability distribution

#### Contextual quality of tag-resource association:

- Value-added: information not in the existing content or metadata
- Relevance and uniqueness: TF/IDF (term frequency / inverse document frequency), user feedback/ratings, usage statistics
- Timeliness: recency and frequency of tag-resource associations

Note that the quality dimensions and metrics may be different for different tagging systems. The social utility of tags certainly varies across systems with different purposes, such as between academic paper databases and fashion apparel catalogs.

## EMPIRICAL ANALYSIS AND EVALUATION OF THE METHODOLOGY

Since tagging systems are relatively new and the tags in these systems continuously evolve, we have not developed a good understanding of tag properties. By observing and analyzing the properties of the tags in several systems, we can identify the tag properties related to tag quality. Such findings will be used as useful feedback for adapting and improving the tag quality assessment methodology.

### Data Collection

We have collected tags from several tagging systems and are in the process of obtaining additional tags from other systems. The tagging systems to be used in the empirical study include:

- Bibsonomy: The system has a periodic database dump of the tags. We have performed initial analysis of the data as of Jan 2008, which contained 58,755 tags.
- CiteULike: The system has daily database dump. We have performed initial analysis of the data as of Jan 2008, which contained 238,785 tags.
- Connotea: The system offers Web API for retrieving the tag data and a library (in Java) to process the data. We have obtained tags from more than 377,000 posts to Connotea collected in December 2007.
- Delicious: Although the system does not offer an API to download tag data, we plan to use and extend a Python-based API ([http://www.michael-noll.com/wiki/Del.icio.us\\_Python\\_API](http://www.michael-noll.com/wiki/Del.icio.us_Python_API)) that uses mining and screen-scraping techniques to download tags.
- The latest ACM classification thesaurus (<http://www.acm.org/about/class/>), a controlled vocabulary with 1207 classes, which may be used for comparison and evaluation purposes when we analyze the tags for documents that are available from the ACM digital library.

In addition, we also plan to download tags programmatically from Flickr site. It will be interesting to compare the tag properties between systems that have different types of resources (textual resources in bookmarking systems and digital photos in Flickr).

### Data Analysis

We will employ data/text mining and natural language processing techniques to analyze the tags. The analyses will include:

- The frequency of typos, acronyms, and compound words
- The frequency of various symbols used for combining multiple words (e.g.,  $\_$ ,  $-$ ,  $!$ , etc.)
- The frequency of polynoms, synonyms, and homonyms
- Statistics on tag length, size of tag set and tag bag applied to a resource, and size of tag clouds
- The similarity (conversely, the distance) between tag set of a resource and the existing metadata of the resource
- Classification of tags according to parts of speech (e.g., noun, verb, etc.)
- Inter-tag relationship analysis, such as subterms and hierarchical relationship between tags with different levels of specificity (e.g., animal vs. cat).
- Convergence of tags

The lexical database WordNet (Fellbaum 1998) will be used in certain linguistics related analyses. For larger tag collections such as those in the millions in tag count, we may use sampling techniques (Ballou et al. 2006) to reduce the dataset for data analysis.

### CONCLUSION AND FUTURE RESEARCH

Despite various discussions and cursory observations about tag quality, there has been little work done to systematically investigate the tag quality issue of tagging systems. We have proposed a methodology for assessing tag quality in three aspects. By the time of conference, we anticipate that we can complete most of the planned empirical analyses. The findings should contribute to our understanding of tag properties and to the enhancement of the proposed methodology.

As future research, we will study the “system quality” of tagging systems in addition to “data quality” of tags, as the system quality clearly affects the data quality. We will build on the findings of the empirical analyses to develop mechanisms for inducing contribution of high quality tags. The mechanisms may include intelligent suggestions of tags to eliminate typos, merge variations of the same tag, and expand tag inclusions based on network and semantic analysis. The mechanisms may also include ways to encourage usage of high-quality tags, such as by tag cloud algorithms that present tags according to their quality as well as popularity. These mechanisms will be implemented in a prototype tagging system, which will be used to evaluate the effectiveness of these quality-enhancing mechanisms.

### REFERENCES

1. Ames, M., and Naaman, M. 2007. "Why We Tag: Motivations for Annotation in Mobile and Online Media," in: *ACM SIGCHI*, B. Begole and S. Payne (eds.). ACM, pp. 971-980.

2. Ballou, D.P., Chengalur-Smith, I.N., and Wang, R.Y. 2006. "Sample-Based Quality Estimation of Query Results in Relational Database Environments," *IEEE Transactions on Knowledge and Data Engineering* (18:5), pp 639-650.
3. Doctorow, C. 2001. "Metacrap: Putting the Torch to Seven Straw-Men of the Meta-Utopia." Retrieved March 1, 2009, 2009, from <http://www.well.com/~doctorow/metacrap.htm>
4. Fellbaum, C. (ed.) 1998. *Wordnet: An Electronic Lexical Database*. Cambridge: The MIT Press.
5. Golder, S.A., and Huberman, B.A. 2006. "Usage Patterns of Collaborative Tagging Systems," *Journal of Information Science* (32:2), pp 198-208.
6. Guy, M., Powell, A., and Day, M. 2004. "Improving the Quality of Metadata in Eprint Archives," *Ariadne* (38:1).
7. Guy, M., and Tonkin, E. 2006. "Tidying up Tags?," *D-Lib Magazine* (12:1).
8. Hammond, T., Hannay, T., Lund, B., and Scott, J. 2005. "Social Bookmarking Tools (I): A General Review," *D-Lib Magazine* (11:4).
9. Kuchinskas, S. 2005. "Flickr to Add Print to Photo Service." Retrieved March 1, 2009, 2009, from <http://www.internetnews.com/ec-news/article.php/3512866>
10. Lee, Y.W., Strong, D.M., Kahn, B.K., and Wang, R.Y. 2002. "Aimq: A Methodology for Information Quality Assessment," *Information and Management* (30:2), pp 133-146.
11. Marlow, C., Naaman, M., boyd, d., and Davis, M. 2006. "Ht06, Tagging Paper, Taxonomy, Flickr, Academic Article, to Read," in: *17th Conference on Hypertext and Hypermedia (HYPERTEXT'06)*. pp. 31-40.
12. Mathes, A. 2004. "Folksonomies - Cooperative Classification and Communication through Shared Metadata." Retrieved March 1, 2009, 2009, from <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>
13. Merholz, P. 2004. "Metadata for the Masses." Retrieved March 1, 2009, 2009, from <http://www.adaptivepath.com/ideas/essays/archives/000361.php>
14. Pipino, L.L., Lee, Y.W., and Wang, R.Y. 2002. "Data Quality Assessment," *Communications of the ACM* (45:4), pp 211-218.
15. Shirky, C. 2005. "Ontology Is Overrated: Categories, Links, and Tags " Retrieved March 1, 2009, 2009, from [http://shirky.com/writings/ontology\\_overrated.html](http://shirky.com/writings/ontology_overrated.html)
16. Sinclair, J., and Cardew-Hall, M. 2008. "The Folksonomy Tag Cloud: When Is It Useful?," *Journal of Information Science* (34:1), pp 15-29.
17. Stivia, B., and Gasser, L. 2008. "Value Based Metadata Quality Assessment.," *Library & Information Science Research* (30:1), pp 67-74.
18. Taylor, A.G. 2003. *The Organization of Information*, (2nd ed.). Westport: Libraries Unlimited.
19. Wang, R.Y., and Strong, D.M. 1996. "Beyond Accuracy: What Data Quality Means to Data Consumers," *Journal of Management Information Systems* (12:4), pp 5-34.
20. Wu, H., Zubair, M., and Maly, K. 2006. "Harvesting Social Knowledge from Folksonomies," *HYPERTEXT '06: Proceedings of the seventeenth conference on Hypertext and hypermedia*, Odense, Denmark, pp. 111-114.